

令和 2 年 6 月 11 日現在

機関番号：13901

研究種目：基盤研究(B) (一般)

研究期間：2017～2019

課題番号：17H01763

研究課題名(和文) 空気/体内伝導音信号の相補的活用に基づく音声/音環境情報処理基盤の構築

研究課題名(英文) Development of fundamental technology for speech and sound event processing based on complementary use of air- and body-conducted sound signals

研究代表者

戸田 智基 (Toda, Tomoki)

名古屋大学・情報基盤センター・教授

研究者番号：90403328

交付決定額(研究期間全体)：(直接経費) 13,700,000円

研究成果の概要(和文)：本研究では、物理的な制約を超えて音信号に含まれる情報の利活用を可能とする技術の実現を目指し、空気/体内伝導音信号の相補的活用に基づく音声/音環境情報処理基盤の構築に取り組んだ。空気/体内伝導音信号を同時収録するための基盤技術を構築するとともに、空気/体内伝導音信号間の対応関係を統計的にモデル化することで、各信号が各々別種の音源信号を優位に含むという特性を活用した空気/体内伝導音信号処理基盤技術を構築した。さらに、発声音声/聴取音強調および発声音声/聴取音認識を実現するための基盤技術を構築し、身体機能拡張への応用にも取り組んだ。

研究成果の学術的意義や社会的意義

空気伝導音信号を対象とした音声/音環境情報処理技術が盛んに研究されている状況の中、本研究では、体内伝導音信号の利活用という別の視点から、新たな音声/音環境情報処理基盤の構築に取り組んだ。空気/体内伝導音信号の相補的活用と深層学習に代表される最先端の機械学習を組み合わせることで、音の重ね合わせによる情報消失といった本質的な問題を緩和できることを学術的に示した。また、本基盤技術を応用することで、身体的機能拡張といった社会的意義の高い応用技術が実現できる可能性を見出した。

研究成果の概要(英文)：In this research, we developed fundamental technology for speech and sound event processing based on complementary use of air- and body-conducted sound signals to make it possible to handle various information included in sound signals beyond physical constraints. We developed fundamental technology to simultaneously record air- and body-conducted sound signals and air- and body-conducted sound signal processing technology capable of effectively using complementary properties of these two types of sound signals. Furthermore, we developed fundamental technology for speech and sound source enhancement processing and speech and sound event recognition processing, further investigating their potential to develop applications for augmenting our physical functions.

研究分野：音メディア情報処理

キーワード：音声情報処理 音響信号処理 音声変換 音声強調 音声認識 音響イベント検出

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 声、音楽、環境音など、日々の生活は実に様々な音に囲まれている。我々は、個々の音が重なり合った空気伝導音信号を聴取することで、例えば、その場の環境を把握し、音源の種類や位置を推測し、音声に至っては埋め込まれた意図や話者情報までも理解するなど、そこに含まれる多種多様な情報を処理する能力を備えている。一方で、物理的な性質上、空気伝導音信号は時間とともに減衰し消失するため、そこに含まれる情報を活用するためには、その場で得た情報を記憶するなり文字媒体などで書き留めるなりする必要がある。時間的な制約を強く受けるため、記録できる情報量は限定され、また、我々の音情報処理能力にも限界があるため、情報の取捨選択が必須となり、必要な情報を漏れなく得ることは困難となる。身体的機能の衰えや障害を患うと、この問題はより一層深刻化し、最低限必要な情報を得ること自体も困難となる。

(2) 時間的な制約や我々の能力的・身体的な制約を超えて、空気伝導音信号に含まれる情報を活用するためには、高度な音情報処理の実現が必要不可欠である。その際に、本質的に問題となるのが、複数音の重ね合わせにより所望の音情報が消失するといった空気振動の物理的性質である。空間的な情報や音源信号の統計的性質(スパース性や独立性)を活用する技術の進展により、音情報処理の性能は着実に改善しているものの、重ね合わせによる情報消失の影響は大きく、処理性能には一定の限界がある。

(3) 空気伝導音信号が本質的に抱える問題への対処として、音声情報処理の研究分野において、体内伝導音信号の活用が検討されている。体内伝導音信号は、主に自身が発声した自己発声音信号を優位に含み、外部音信号は大幅に低減されるという性質を持ち、また、周囲の人が聴取困難なほど微かな音声(非可聴つぶやき)も収録可能であるという特徴も有する。しかしながら、体内伝導収録の影響により、著しい音質劣化が生じる。これに対し、空気伝導音信号は、音質劣化を生じさせることなく、自己発声音信号や外部音信号など多種多様な音信号の収録が可能である。しかしながら、多数の音信号の重ね合わせによる情報消失の影響を強く受ける。これらの空気/体内伝導音信号の性質を相補的に活用することで、重ね合わせによる情報消失の影響を緩和し、これまでの空気伝導音信号のみを対象としてきた音情報処理では困難であった機能が実現できると期待される。

2. 研究の目的

(1) 本研究では、物理的な制約を超えて、音信号に含まれる情報の利用を可能とする技術の実現を目指し、空気/体内伝導音信号の相補的活用に基づく音声/音環境情報処理基盤を構築することを目的とした。空気伝導音信号と体内伝導音信号が各々別種の音信号を優位に含むという性質を活用することで、複数音信号の重ね合わせにより生じる所望の音情報の消失といった空気伝導音信号が持つ本質的問題の解決を図った。具体的には、空気/体内伝導音信号の収録基盤技術の構築、空気/体内伝導音信号間の対応関係をモデル化することで相補的活用を可能とする空気/体内伝導音信号処理基盤技術の構築、さらには、発声音声/聴取音強調基盤技術ならびに発声音声/聴取音認識基盤技術の構築に取り組んだ。

3. 研究の方法

(1) 空気伝導音声信号と体内伝導音声信号間の対応関係を統計的にモデル化する技術を構築し、各種音声情報処理および環境音情報処理との統合を図るとともに、身体機能拡張への応用に向けた検討を行った。具体的には、空気/体内伝導音信号の相補的活用に基づく音声/音環境情報処理基盤技術として、以下の3つの研究課題に取り組んだ。

(2) **課題1「空気/体内伝導音信号に対する収録・処理基盤技術の構築」**: 空気伝導音信号と体内伝導音信号を相補的に活用した音声/音環境情報処理の研究を実施する上で、主に以下の2つの部分課題に取り組んだ。

課題1A「空気/体内伝導音の収録基盤技術の構築」: 既存の空気伝導マイクロフォンおよび体内伝導マイクロフォンを使用して、ウェアラブル型の多チャンネル空気・体内伝導音収録装置の試作に取り組んだ。また、実際に試作した収録装置を用いて、空気/体内伝導音信号の同時収録を行い、実験用データセットの構築に取り組んだ。

課題1B「空気/体内伝導音に対する信号処理基盤技術の構築」: 空気伝導音声信号と体内伝導音声信号間の対応関係を利用して複数音信号の重ね合わせの問題を緩和するために、空気/体内伝導音信号を処理する基盤技術の構築に取り組んだ。具体的には、ウェアラブル型空気/体内伝導音収録装置を装着した話者が発声した自己発声音声を抽出する自己発声音声強調技術と、自己発声音声を抑圧して他の音を抽出する自己発声音声抑圧技術の構築に取り組んだ。

(3) **課題2「発声音声および聴取音に対する強調基盤技術の構築」**: 発声機能および聴覚機能の拡張といった応用技術への発展を視野に入れて、主に以下の2つの部分課題に取り組んだ。

課題2A「発声音声強調基盤技術の構築」: 発声音声を強調する処理として、統計的音声変換処理に着目し、深層学習や音声生成過程の物理的仕組みを導入することで、音質改善およ

び柔軟性改善に取り組んだ。音声波形生成基盤技術として、音声波形加工モデルや深層音声波形生成モデルの構築に取り組むとともに、音声特徴量変換基盤技術として、深層系列変換に基づく音声特徴量系列変換法や教師なし音声変換モデル学習法の構築に取り組んだ。

課題 2 B 「聴取音強調基盤技術の構築」: 各種音源信号を強調する処理として、音源分離処理に着目し、各音源信号の統計的性質を事前情報として活用する枠組みの実現に取り組んだ。統計的音源モデルと物理的な空間モデルを併用した音源分離の枠組みに対して、深層学習に基づく音源信号モデルの導入や、より汎用的な音源分離問題への拡張に取り組むとともに、音声信号のみならず音楽信号のような様々な音源信号を対象とした強調技術の構築に取り組んだ。



図 1 .ウェアラブル型多チャンネル空気 / 体内伝導音収録装置

(4) **課題 3 「発声音声および聴取音に対する認識基盤技術の構築」:** 記憶機能の拡張といった応用技術への発展を見据えて、主に以下の 2 つの部分課題に取り組んだ。

課題 3 A 「発声音声認識基盤技術の構築」 秘匿性の高いサイレント音声インタフェースの実現に向けて、体内伝導音声として非可聴つぶやきを用いた発声音声認識処理に着目し、深層学習の導入や、発話者および音環境への適応技術、さらには、音声コマンドによる情報操作に向けた認識技術の構築に取り組んだ。また、感情などのパラ言語情報を対象とした発声音声認識処理にも着目し、教師ラベル無し音声データを有効活用する技術の構築に取り組んだ。

課題 3 B 「聴取音認識基盤技術の構築」 発話内容に加えて音環境の自動記録の実現を視野に入れて、環境音を対象とする認識技術の構築に取り組んだ。特に、様々な音イベントに対して共有可能な記号表現を獲得する技術の構築や、深層学習に基づく音イベント検出技術、ならびに、異常音検出技術の構築に取り組んだ。

4 . 研究成果

(1) 3 つの研究課題「空気 / 体内伝導音信号に対する収録・処理基盤技術の構築」、「発声音声および聴取音に対する強調基盤技術の構築」、「発声音声および聴取音に対する認識基盤技術の構築」の各々における個々の部分課題に対して、以下の研究成果が得られた。

(2) **成果 1 A 「空気 / 体内伝導音の収録基盤技術の構築」:** 体内伝導マイクロフォンとしてネックバンド型非可聴つぶやきマイクロフォンを利用するとともに、ネックバンドに空気伝導マイクロフォンを取り付けることで、発声者の襟元に装着可能なウェアラブル型の多チャンネル空気 / 体内伝導音収録装置を試作した (図 1 参照)。これを用いて、多チャンネル空気伝導音信号と体内伝導音信号の同期収録を行い、発声者の自己発聲音声の収録を行った。また、発声者の四方八方から提示される環境雑音の収録も行った。これらの収録音を用いて、実験用空気 / 体内伝導音信号データセットを構築するとともに、本研究課題の実験的評価において利用した。

(3) **成果 1 B 「空気 / 体内伝導音に対する信号処理基盤技術の構築」:** 統計的音源モデルと物理的制約を踏まえた空間モデルを併用した多チャンネル音源分離手法を発展させることで、空気伝導音信号と体内伝導音信号を相補的に活用した自己発聲音声強調技術、および、自己発聲音声抑圧基盤技術を構築した。

音源モデルとして、空気伝導音声信号と体内伝導音声信号間の非線形な対応関係を捉えるために、非負値行列因子分解 (Non-negative matrix factorization: NMF) における基底ベクトルを拡張することで、空気 / 体内伝導音信号の結合確率密度モデリングを実現する手法を考案した。これにより、多チャンネル空気伝導音声信号に対する自己発聲音声強調・抑圧用線形分離フィルタの推定において、空気伝導音信号のみでなく、体内伝導音信号も併用した処理を可能とし、強調・抑圧性能を改善できることを示した (図 2 参照)。

さらなる性能改善に向け、非線形分離フィルタの利用も可能とする劣決定音源分離の枠組みへの拡張を行うとともに、変分自己符号化器 (Variational autoencoder: VAE) に基づく深層音源モデルに基づく空気 / 体内伝導音声の結合確率密度モデリングを実現する手法を考案した。これにより、自己発聲音声強調・抑圧処理の性能を大幅に改善できることを示した (図 2 参照)。

これらの研究成果を査読付き国際会議論文などで発表した (例えば、 5 . 主な発表論文等 [学会発表] [Takada 他, EUSIPCO 2020](#) および [Takada 他, APSIPA ASC 2018](#) など)。

(4) **成果 2 A 「発聲音声強調基盤技術の構築」:** 統計的音声変換に基づく発聲音声強調基盤技術として、深層学習に基づく高品質かつ操作性の高い音声波形生成基盤技術、ならびに、高品質かつ柔軟性の高い音声特徴量変換基盤技術を構築した。

音声波形生成基盤技術として、音声波形直接加工処理の枠組みに着目し、耐雑音性を捉える客観指標を考案するとともに、音高を保持した声質変換においては外部雑音に頑健かつ高品質な音声変換処理が可能であることを示した。また、音高変換も可能とする枠組みにおいても高品質な音声変換処理を実現するために、畳み込みニューラルネットワークや敵対的生成ネットワークなどの深層学習モデルを音声波形生成処理に活用する枠組みを考案した。さらに、音声生成過程の物理的な仕組みを反映させる

	従来法	提案法 1	提案法 2	提案法 3
空間モデル	ランク1制約 線形分離	ランク1制約 線形分離	制約なし 非線形分離	制約なし 非線形分離
音源モデル	NMF	結合NMF	結合NMF	結合VAE

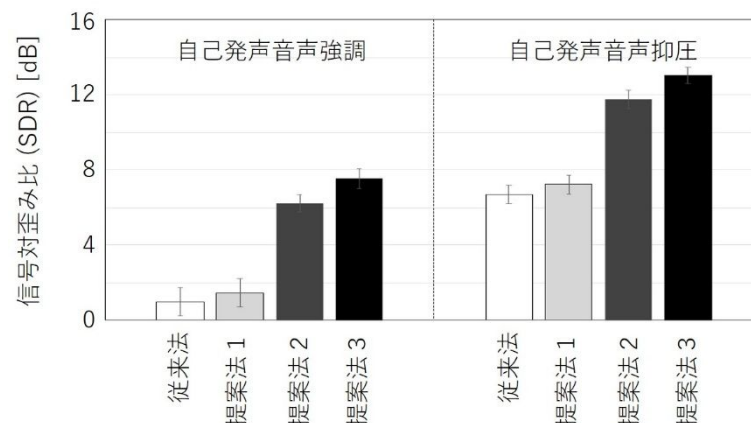


図2 . 自己発声音声強調・抑圧性能の比較

ために、従来の音声波形生成モデルとの併用を可能とする深層音声波形加工ネットワークや、適応型拡張畳み込みニューラルネットワークを考案し、高い操作性を保持したまま高品質な音声波形の合成処理を可能とする深層音声波形生成基盤技術を構築し、その有効性を示した。音声特徴量変換基盤技術として、深層系列変換モデルに基づく音声特徴量系列変換手法を考案し、声質のみでなく話し方などの変換も可能とするとともに、外国語発音補正処理といった機能拡張への応用に適用し、その有効性を示した。また、VAEなどの深層学習モデルを導入することで、同一発話対を必要としない教師なし音声変換モデル学習手法を考案し、変換モデル学習時の制約を緩和することで、音声変換技術の応用範囲を拡大した。

これらの研究成果を査読付き学術論文および国際会議論文などで発表した(例えば、5 . 主な発表論文等 [雑誌論文] [Kameoka 他, IEEE/ACM TASLP, 2019](#)、[学会発表] [Wu 他, INTERSPEECH 2019](#) および [Tanaka 他, ICASSP 2019](#) など)。また、国際会議にてチュートリアル講演や招待講演を行った(5 . 主な発表論文等 [学会発表] [Toda 他, INTERSPEECH 2019](#) および [Kameoka 他, SANE 2019](#))。

(5) 成果2 B 「聴取音強調基盤技術の構築」: 多チャンネル音源分離に基づく聴取音強調基盤技術として、統計的音源モデルと空間モデルを併用する枠組みに対して、VAE に基づく深層音源モデリング技術を導入し、分離性能の向上を達成した。また、音源分離と残響抑圧を同時に最適化する技術も構築し、その有効性を示した。さらに、観測チャンネル数(マイクロフォン数)よりも音源数が多いといった劣決定音源分離問題への拡張を行うとともに、音楽を含む多種多様な音源を対象とした音源分離処理を実現するなど、より汎用的な条件下においても動作する音源分離基盤技術を構築した。これらの研究成果を査読付き学術論文および国際会議論文などで発表した(例えば、5 . 主な発表論文等 [雑誌論文] [Seki 他, IEEE Access, 2019](#) および [Kameoka 他, Neural Computation, 2019](#)、[学会発表] [Inoue 他, ICASSP 2019](#) など)。

(6) 成果3 A 「発声音声認識基盤技術の構築」: 発声音声認識基盤技術として、体内伝導音声である非可聴つぶやきに対する認識処理に対し、深層学習に基づく発話者および音環境への適応技術を導入することで、秘匿性の高いサイレント音声入力を実現する基盤技術を構築した。また、深層学習に基づく音声コマンド認識処理の軽量化、ならびに、表現学習に基づく感情音声認識処理を実現し、音声入力機能の利便性を改善した。これらの研究成果を査読付き国際会議論文などで発表した(例えば、5 . 主な発表論文等 [学会発表] [Yamamoto 他, INTERSPEECH 2019](#) 参照など)。

(7) 成果3 B 「聴取音認識基盤技術の構築」: 聴取音認識基盤技術として、環境音の記号化処理を可能とするために、記号表現として擬音語表現に着目し、深層学習に基づく音信号から擬音語表現への変換技術を構築した。また、音響情報とテキスト情報を結び付ける特徴空間の獲得手法を考案し、その有効性を示した。さらに、個々の音イベントの検出や異常音の検出を可能とする深層ニューラルネットワークなど、深層学習に基づく音環境情報処理基盤技術を構築した。この他にも、体内伝導嚙下音を活用した嚙下音障害診断支援技術も構築した。これらの研究成果を査読付き学術論文および国際会議論文などで発表した(例えば、5 . 主な発表論文等 [雑誌論文] [Miyazaki 他, IEEJ Trans., 2019](#) および [西村 他, 日本知能情報フアジィ学会誌, 2018](#)、[学会発表] [Miyazaki 他, EUSIPCO 2018](#) など)。

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 8件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 Shogo Seki, Hirokazu Kameoka, Li Li, Tomoki Toda, Kazuya Takeda	4. 巻 Vol. 7, No. 1
2. 論文標題 Underdetermined source separation based on generalized multichannel variational autoencoder	5. 発行年 2019年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 168104-168115
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2019.2954120	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Hirokazu Kameoka, Li Li, Shota Inoue, Shoji Makino	4. 巻 Vol. 31, No. 9
2. 論文標題 Supervised determined source separation with multichannel variational autoencoder	5. 発行年 2019年
3. 雑誌名 Neural Computation	6. 最初と最後の頁 1891-1914
掲載論文のDOI（デジタルオブジェクト識別子） 10.1162/neco_a_01217	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hirokazu Kameoka, Takuhiro Kaneko, Kou Tanaka, Nobukatsu Hojo	4. 巻 Vol. 27, No. 9
2. 論文標題 ACVAE-VC: non-parallel voice conversion with auxiliary classifier variational autoencoder	5. 発行年 2019年
3. 雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	6. 最初と最後の頁 1432-1443
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2019.2917232	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Koichi Miyazaki, Tomoki Toda, Tomoki Hayashi, Kazuya Takeda	4. 巻 Vol. 14, No. 3
2. 論文標題 Environmental sound processing and its applications	5. 発行年 2019年
3. 雑誌名 IEEJ Transactions on Electronics, Information and Systems	6. 最初と最後の頁 340-351
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/tee.22868	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 西村良太, 檜垣美帆, 北岡教英	4. 巻 Vol. 30, No. 4
2. 論文標題 RNN-LSTMによる音響ベクトル空間と文書ベクトル空間とのマッピング	5. 発行年 2018年
3. 雑誌名 日本知能情報ファジィ学会誌	6. 最初と最後の頁 628-633
掲載論文のDOI (デジタルオブジェクト識別子) 10.3156/jsoft.30.4_628	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計89件(うち招待講演 12件/うち国際学会 41件)

1. 発表者名 Moe Takada, Shogo Seki, Patrick Lumban Tobing, Tomoki Toda
2. 発表標題 Semi-supervised enhancement and suppression of self-produced speech using correspondence between air- and body-conducted signals
3. 学会等名 The 28th European Signal Processing Conference (EUSIPCO 2020) (国際学会)
4. 発表年 2021年

1. 発表者名 Shota Inoue, Hirokazu Kameoka, Li Li, Shogo Seki, Shoji Makino
2. 発表標題 Joint separation and dereverberation of reverberant mixtures with multichannel variational autoencoder
3. 学会等名 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Kou Tanaka, Hirokazu Kameoka, Takuhiro Kaneko, Nobukatsu Hojo
2. 発表標題 AttS2S-VC: Sequence-to-sequence voice conversion with attention and context preservation mechanisms
3. 学会等名 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Tomoki Toda, Kazuhiro Kobayashi, Tomoki Hayashi
2. 発表標題 Tutorial: Statistical voice conversion with direct waveform modeling
3. 学会等名 The 20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 YiChiao Wu, Tomoki Hayashi, Patrick Lumban Tobing, Kazuhiro Kobayashi, Tomoki Toda
2. 発表標題 Quasi-periodic WaveNet vocoder: a pitch dependent dilated convolution model for parametric speech generation
3. 学会等名 The 20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Taiki Yamamoto, Ryota Nishimura, Masayuki Misaki, Norihide Kitaoka
2. 発表標題 Small-footprint magic word detection method using convolutional LSTM neural network
3. 学会等名 The 20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Hirokazu Kameoka, Takuhiro Kaneko, Kou Tanaka, Nobukatsu Hojo
2. 発表標題 Voice conversion with image-to-image translation and sequence-to-sequence learning approaches
3. 学会等名 SANE 2019 - Speech and Audio in the Northeast (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Koichi Miyazaki, Tomoki Hayashi, Tomoki Toda, Kazuya Takeda
2. 発表標題 Connectionist temporal classification-based sound event encoder for converting sound events into onomatopoeia representations
3. 学会等名 The 2018 European Signal Processing Conference (EUSIPCO 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Moe Takada, Shogo Seki, Tomoki Toda
2. 発表標題 Self-produced speech enhancement and suppression method using air- and body-conductive microphones
3. 学会等名 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference 2018 (APSIPA ASC 2018) (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	亀岡 弘和 (Kameoka Hirokazu) (20466402)	日本電信電話株式会社NTTコミュニケーション科学基礎研究所・メディア情報研究部・主任研究員 (94305)	
研究 分担者	北岡 教英 (Kitaoka Norihide) (10333501)	豊橋技術科学大学・工学(系)研究科(研究院)・教授 (13904)	